Neural Computing and Applications
https://doi.org/10.1007/s00521-020-05487-1

S.I. : WORLDCIST'20

A novel emergency situation awareness machine learning approach to assess flood disaster risk based on Chinese Weibo

Hua Bai[1] · Hualong Yu[2] · Guang Yu[2] · Xing Huang[3]

Received: 17 July 2020 / Accepted: 27 October 2020

Abstract

Social media emerged as an important resource of information to improve the emergency situation awareness of flooding disasters. However, the online microblog text stream is unstructured and unbalanced obviously. Given the big, real-time, and noisy flood disaster microblog text flow, a new regional emergency situation awareness model to automatic assess flood disaster risk is proposed. Firstly, according to the established online disaster event-meta frame, a multi-label classification algorithm for the flood microbloggings is constructed based on the historical dataset. This algorithm helps to assign the relevant event-meta tags to each situation microbloggings. Second, a new machine learning method for dynamic assessment of flood risk for online microbloggings is developed. The flood event-metas are considered to be feature vectors, and the four different levels of flood risk are considered to be four classes. Then, the flood risk assessment task is innovatively transformed into a multi-classification task. By the logistic regression ordered multi-classification algorithm, the dynamic quantitative evaluation of event-meta, users and regional risks is realized. Finally, the proposed model is applied in the case of the Yuyao Flood. The results of the case study show that the Yuyao Flood's online quantitative risk assessment results are consistent with real accumulated precipitation data, which illustrate that the proposed machine learning model could realize the bottom-up automatic disaster information collecting by processing victim user-generated content effectively. Social media is proven to supplement the deficiencies of traditional disaster statistics and provide real-time, scientific information support for the implementation of flood emergency processes.

Keywords Flood risk assessment · Social media · Emergency situation awareness · Machine learning

1 Introduction

Globally, flood disasters affect more people than any other natural disaster. Flood disasters have the typical characteristics of strong suddenness, wide affected areas, and large casualties and property losses. After entering the twenty-first century, as global warming strengthens and the urbanization process accelerates, the frequency and impact of flood disasters increase. For example, the Yuyao Flood (2013, Zhejiang, China) and the Shouguang Flood (2018, Shandong, China) caused huge material losses and casualties. Therefore, dealing with these sudden flood disasters effectively and reduce disaster losses considerably has important theoretical and practical significance.

The biggest challenge in the emergency management process is communicating information [1]. Flood disaster emergency response puts forward high requirements for the timeliness and accuracy of emergency information processing. After the flood disaster, timely and reliable disaster information plays a key role in rescue decision-making, which can effectively reduce the physical damage and casualties of floods [2]. From the Japanese tsunami (2011) to the floods in Louisiana (2016), many flood disaster events have confirmed the importance of advanced information and communication technologies for disaster emergency management [3].

✉ Xing Huang
xingh@swust.edu.cn

1 School of Economics, Fujian Normal University, Fuzhou, China

2 School of Management, Harbin Institute of Technology, Harbin, China

3 School of Economics and Management, Southwest University of Science and Technology, Mianyang, China

Published online: 21 November 2020

Considering the importance of disaster information, many countries, cities, and research institutions around the world have built a series of disaster information databases at regional, national, and global levels, such as NatCat, Sima, Desinventar, and EM-DAT. However, most of these databases focus on the integration of historical data and lack the collection and analysis of online real-time data generated by experienced users. Given the series of information asymmetry problems in the emergency rescue process of flood disasters, studying the useful theory and method of auto collection and analyzing the situation awareness of real-time flood disaster information is imperative.

After the coming of the Internet, several web-based disaster event monitoring systems (e.g., "Did you feel it?") were successively developed to utilize the large amount of network information resources effectively. These systems perform disaster monitoring and information notifying in real-time by integrating official information, media news, and targeted feedback of website users.

What features flood disaster response in this novel digital age? New ICT technologies that serve flood disaster emergency reduction have great potential function, which have enabled the public to conduct and share a large number of flood observations. In an increasing number of post-flooding situations, citizens use social media tools to post and trans information in order to carrying out emergency tasks and to be potential complement governmental emergency. In post-flood disaster situation, data about the hazard and its consequences are important and difficult to collect. The disaster situation information provided by social media witnesses is an important useful data source and should be explored effectively. With the widespread use of social media tools, they have gradually expanded from daily chat tools to key platforms for individuals and organizations to communicate disaster information [4, 5]. Social media, which is a typical user-generated content communication tool, has the advantages of timeliness and interactivity. Social media can give full play to the collective wisdom and corporate social responsibility in the disaster relief process serving the disaster emergency response process [6, 7]. The important role of social media tools in the disaster process has attracted increasing attention from researchers, which has also been involved in the emergency management of many disasters management practices, such as the Japan earthquake and tsunami (2011), the Queensland flood (2010), and the Victoria flood (2010). However, previous social media studies in flood scenarios are mostly case studies or qualitative description analyses of online disaster information. A few quantitative studies can effectively use online information to support and assist in the flood emergency response process. Determining how to transform the massive, disordered, and unstructured first-hand qualitative description of social media texts of flooding disasters released by users in disaster-affected areas into reasonable, efficient, and accurate quantitative evaluation and analysis results in a short period to provide effective information support for emergency response is urgently needed.

In order to efficiently use social media tools in flood disaster emergency response, a large amount of real-time social media messages needs to be collected in time, and the collected incremental data sets also need to be quickly analyzed. However, given the typical characteristics of short texts, such as sparse and noisy, extracting effective disaster information from a large number of microblog text streams is challenging for automated information extraction technology. The extraction method for traditional knowledge engineering information relies heavily on the background knowledge of experts, requires a lot of human resources, and has low adaptability and portability [8]. Therefore, the traditional method is difficult to apply to unstructured multi-dimensional social media information extraction.

Motivated by the above considerations, we introduce in this work machine learning algorithms to convert the problem of online disaster information quantitative analysis problem into a text automatic classification task based on a large-scale training set, which provides new ideas for the study of flood disaster social media. Therefore, the following research questions will be answered in this paper. (1) What type of information can social media tools provide for flood disaster emergency response? (2) How to transform the qualitative text description of the witness into a quantitative data source? (3) How to quickly and efficiently analyze the social media information of witnesses?

To address all of these questions, this paper is divided into five sections. Section 1 is the introduction and motivation to the research. Section 2 presents previous studies. Section 3 describes the proposed methodology in detail. Section 4 presents a case study on the Yuyao Flood and discusses the key findings of the proposed methods and the results. Section 5 concludes the paper and presents directions for future work.

2 Related work

2.1 The importance of social media in flood disaster situation

Researchers have reached a broad consensus on the important role of flood disaster social media information in emergency response and disposal processes. Many research results have proven that online situation information can effectively improve disaster emergency response and treatment effect and provide a lot of useful information for emergency decision-making. New technologies such as social media provide a great deal of potential for engaging citizens and facilitating the and coproduction of emergency situation knowledge [9].The findings of a pilot study that explored the information experiences of people using social media during the flooding of the 2011 Brisbane River informed the development of social media platforms that can provide relevant and accessible information for the general public in the flood disaster event [10].

Research on social media in flooding disasters began with the outbreak detection of disaster events. Flooding disaster information on social media helps in discovering flooding immediately and determining the emergency response time. In 2012, Commonwealth Scientific and Industrial Research Organization developed the Emergency Situation Awareness system, which can use social media information to monitor flood disasters in real-time [11]. Several countries developed a series of online systems, such as "Twicident," "Tweet4act" [12], "CrisisTracker" [13], "Ushahidi platform," and SWIM [14], for the real-time monitoring of social media disaster information. Disasters often cause great damage to traditional communication facilities. However, Internet infrastructure is robust, which is why social media is often considered to be reliable in disaster situations. Disaster victims that seek help and post first-hand situational information via social media are increasing [15]. Many users use social media mobile terminals to publish a lot of information shortly after they experienced a disaster situation. They also tend to use the functions of replying and commenting for information exchange [16]. Moreover, the information transmission speed and bandwidth of the Internet are often significantly greater than traditional communication facilities and are highly suitable for information communication in disaster situations [17]. Therefore, disaster events are often accompanied by a large amount of social media situation information from disaster areas [18].

2.2 The function of social media in flood disaster situation

The Case of the 2013 European Floods in Germany confirmed that Twitter, Facebook (FB), Google Maps and other services were frequently used by affected citizen and volunteers to coordinate help activities among themselves [19]. A qualitative case study of the 2011 Thailand flooding was carried out by an interpretive approach, which confirmed that three main dimensions of empowerment process from social media to community, including structural, psychological, and resource, were helpful to achieve collective participation, shared identification, and collaborative control [20]. Social media users are recipients and creators of information [21]. When a disaster occurs, social media platforms often generate large-scale related tweets from sensitive areas in a short time [22]. Social media could provide massive first-hand documents for the emergency rescue and flooding disaster reduction process [23]. Crooks (2013) [24] confirmed that social media is considerably superior to the web information "Did you feel it?" system in terms of information dissemination speed and capacity. Using social media to monitor the occurrence and development of disaster events may be more effective than traditional methods [25]. Thus, social media tools can provide long escape time for disaster victims and detailed disaster situation information for official rescue [26]. In the emergency response process of flooding disasters, victims tend to search, ask for help, and publish the latest water level and other disaster information at any time via social media, which help in the timely understanding of the latest disaster progress in various places [27]. In the case where obtaining information on the disaster situation in some disaster-stricken areas by traditional means is difficult, the first-hand social media information released by the people in those areas can effectively provide information support for disaster emergency and rescue decision-making [28, 29]. Moreover, the disaster situation information posted via social media has many other functions, such as helping emergency agencies to discover and report public health problems [30], finding missing people, identifying victims' material needs [31], and targeting allocation of medical aid resources [32].During the 2016 flood in Louisiana, Parishes actively used their social media to share disaster information with the flood-affected community, e.g., flood inundation map, emergency shelters' locations, and medical services [33]. A new method is proposed to assess transportation infrastructure damage by collecting remote-sensing imagery from social media during 2013 Colorado floods [34]. Besides, an algorithm to identify victims asking for help during flood disaster was developed using Twitter post in a flood-related disaster. The algorithm worked well and had more than 80% classification accuracy and location prediction accuracy [35]. Mobile phone networks and social media are critical to information seeking in a flood [36].

2.3 Flood risk analysis by social media data

In recent years, the effectiveness and practicability of social media in flood risk assessment have gradually attracted attention. Social media is widely used in disaster information dissemination in countries around the world. The findings of 2010 Pakistan Floods study proposed that there was a slight preference for linking to social media for Pakistani users on Twitter [37]. The results of microblogging text messages study of the Twitter platform (tweets) produced during the River Elbe Flood of June 2013 in Germany showed that messages near to severely flooded areas have a much higher probability of being related to floods [38]. Two natural disasters case studies confirmed that social media helped social media users in Saudi Arabia to communicate floods' damage gravity; discuss responsible; criticize the government; call for action to remedy flood situation; and express sadness over the loss [39]. Three recent citizen science projects of Argentina, France, and New Zealand were launched independently to compute valuable hydraulic data such as the extent and depths of inundated areas and flow rate estimates by citizens' messages, photographs, and videos [40]. A case study that uses data from Twitter to detect and locate flood events in the UK demonstrated that high-resolution social sensing of floods is feasible and high-quality historical and real-time maps of floods using Twitter also could be produced [41].

2.4 Research methods of flood social media

The big data technologies such as machine learning are gradually applied to the emergency information extraction process of social media [42, 43]. Geo-social media data form flood disasters could instead hydrometeorological data for streamflow estimation and flood monitoring [2]. The geographic information carried by flood disaster social media can provide disaster management and rescuers with detailed situation information of the affected area [44], which is convenient for identifying and predicting disaster severity [45]. Neubaum [46] studied the trend of social media data flow in Egypt between 2013 and 2014 and found that the time change of social media activities is consistent with the rainstorm process and that social media activities are significantly positively correlated with precipitation intensity. Fohringer et al.(2015) developed a tool named "PostDistiller" to filter geolocated posts from social media services which include links to photographs in flood disaster [47]. Smith [48] showed that the data flow of flood disasters related to social media platforms has good temporal and spatial synchronization with flood events. Smith presented a real-time modeling framework to identify areas likely to have flooded using data obtained only through social media was presented and proven to be effectively by Twitter data during two 2012 flood events. Li [49] drew a map of the Southern California flood in 2015 based on social media data and showed that social media such as Twitter have emerged as a new data source for disaster management and flood mapping. The results of Brenden's research demonstrated that social media data sources could be used to quickly obtain the understanding of the location, the timing, the causes and impacts of floods [50]. Deep learning technology was used to extract of pluvial flood relevant Volunteered Geographic Information from User Generated microblogs [51, 52].

Currently, social media tools such as Twitter and Facebook have been widely used in flood emergency response researches and practices. Besides, the role of social media tools in flooding situation has also been confirmed in non-English-speaking countries and regions. Liang [53] pointed out that the data from Weibo, which is one of the most popular social media platforms in China, are closely related to the development of typhoon and flooding disasters. Rescuers can accurately grasp the real-time situation of typhoons and floods on a small scale via Weibo [54]. That is, Weibo can serve flood disaster prediction and early warning process effectively. However, compared with Twitter and other English social media tools, quantitative analyses of flood disaster data on Chinese social media platforms are limited [55]. Moreover, previous studies on the mining of emergency functions for flood disaster situation information on social media remain at the post-mortem analysis stage of historical data, and no effective connection is observed between the real-time processing of online data streams and emergency decision response. Therefore, determining how to use Chinese social media effectively to translate the subjective qualitative description of disaster victims into the quantitative analysis required for emergency rescue decision-making and how to achieve quantitative research and rapid response to flood impact are important issues to be solved.

3 Methodology

After the flood disaster, a social media platform can provide a large amount of first-hand disaster information for rescue and disaster reduction decisions promptly. However, determining how to extract the disaster information accurately from the noisy and huge text stream to meet the decision-making needs has become the difficulty and focus of online flood disaster information mining. In flooding disaster situations, emergency decision-makers need to grasp relevant knowledge in a short time, including disaster factor, disaster-bearing body, disaster event status, and the environment.

Therefore, to realize the extraction process of real-time online disaster knowledge, this study first builds an online disaster event-meta frame to transform the quantitative assessment of disaster information into a multi-label classification problem of machine learning and assign disaster information to its associated event-meta class labels. Then, facing the quantitative needs of the decision-making process, this study selects a four-classification machine learning model to assess quickly the event-meta, microblog instance, and regional flood disaster risk levels (i.e., Levels

I–IV) of the flooding risk classification evaluation model based on social media. The above two steps help in the online assessment of flooding risk levels in different affected areas via real-time flood social media data stream. Moreover, the risk level evaluation results of different event-metas can be used to realize the real-time and rapid feedback of flooding disaster information such that online data can well adapt to emergency decision-making needs. The details of the research idea are shown in Fig. 1.

According to the above design ideas, the integration framework of social media flood information established in this study mainly includes the following processes:

(1) The disaster event-meta, which is the online disaster extraction label, is determined based on the automatic clustering results of historical flood social media data and the information requirements of emergency response decision-making.
(2) For online real-time disaster information on social media, incremental multi-label classification is performed based on event-meta class labels to determine the event-meta to which online disaster information belongs.
(3) For flood microblogs posted by individual users, the real-time microblog text is used as input to evaluate the event-meta risk level. The post-evaluation event-meta level is then used as input to conduct a risk level assessment of the users' disaster situations.
(4) Online flood microblogs from a certain area within the time window are clustered. Then, the regional event-meta risk assessment results are calculated according to the mode. Next, the regional disaster risk level is evaluated on the basis of the event-meta risk level input of the region.

The above process is shown in Fig. 2.

3.1 Event-meta analysis of flood social media

An event-meta is an important means to describe event information. The establishment of an event-meta in the field need analysis and professional knowledge constructs recognized concepts in the field. The inner connections of these concepts are described at different levels in structured language to realize the understanding of events in the field.

The online disaster information event-meta is an important part of the classification and extraction of disaster situation information on social media. Generally, the traditional disaster event-meta analyzes the perspective of disaster factor, disaster-bearing body, status information, and environmental information. However, disaster information on social media has typical characteristics, such as colloquialization and fragmentation. Users often use many emoticons, network languages, and punctuation marks to express information. These expressions are different from those in traditional written text. Therefore, a meta-analysis framework for online flood disaster social media information is constructed to meet emergency decision-making needs. According to the meta-analysis structure of disaster events and the consultation results from experts using the Delphi method, a novel flooding disaster event-meta framework is developed, as shown in Fig. 3. The framework is the basis of the real-time knowledge extraction of flood disaster situation. Each event-meta label in the framework is the class label for the next incremental multi-label clustering.

3.2 Multi-label classification algorithm of short text

The above online flooding event-meta framework helps to realize online disaster knowledge extraction. An event-

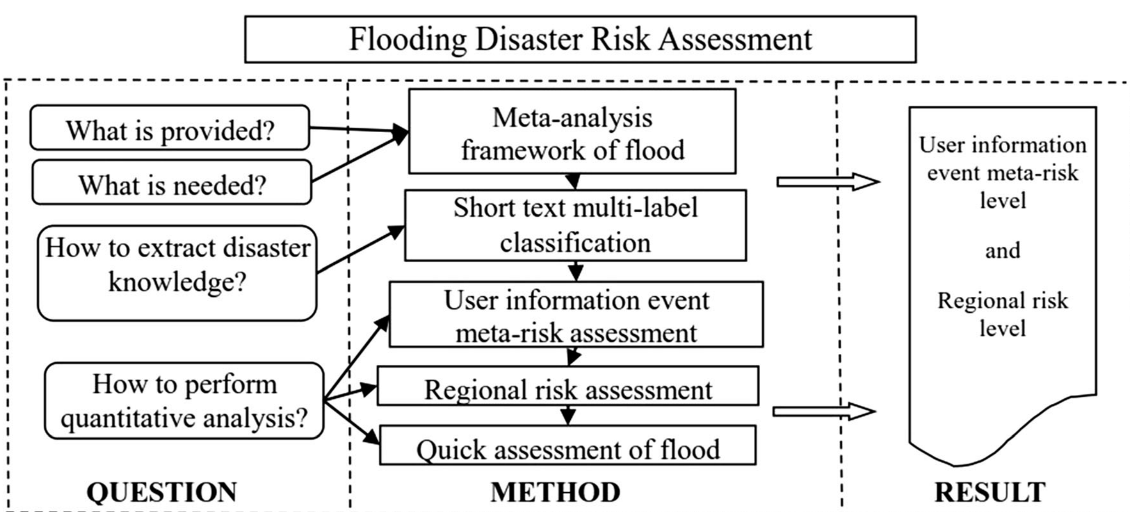


Fig. 1 Main idea of flooding disaster situation awareness

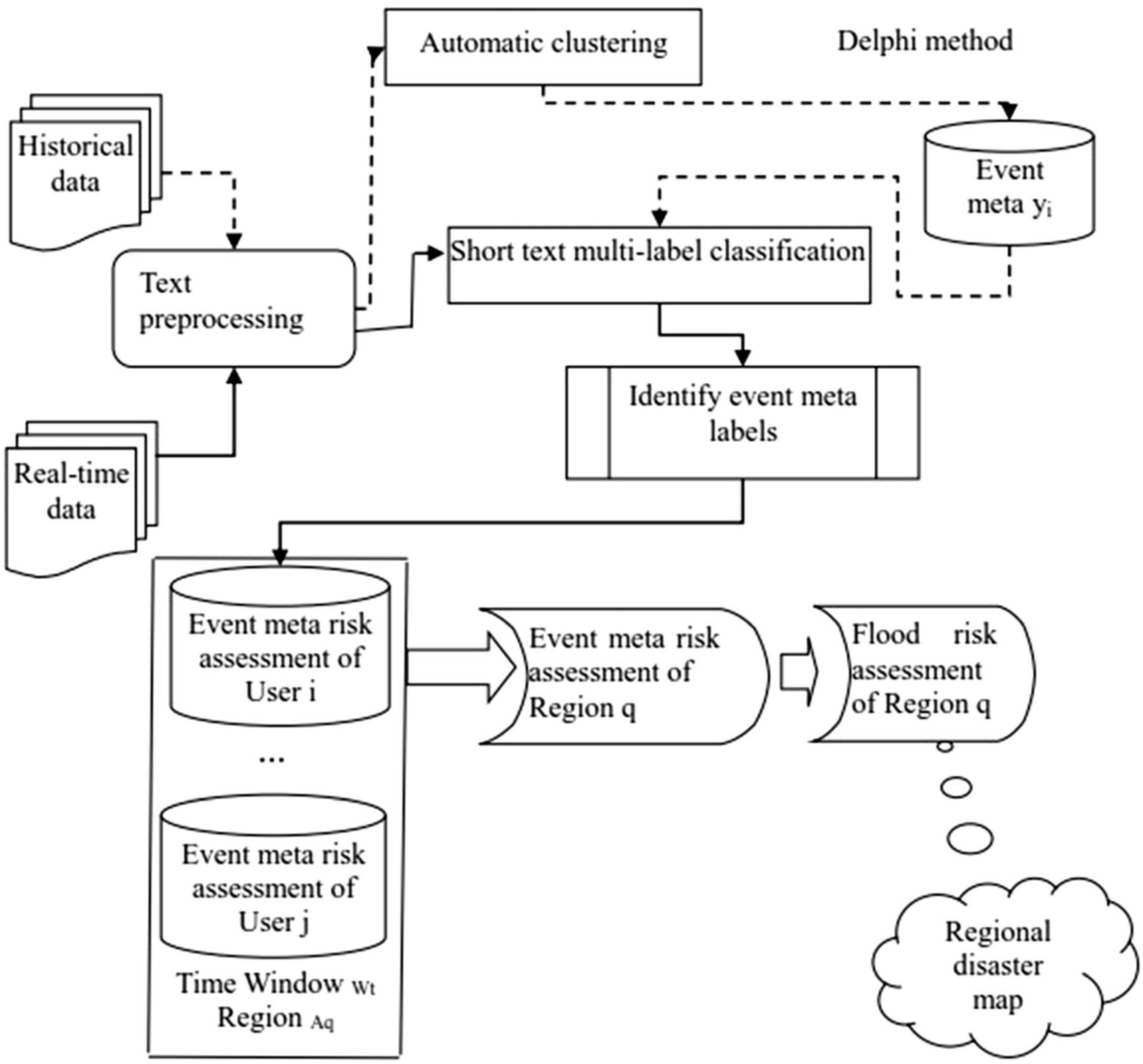


Fig. 2 Process of flooding disaster risk assessment

meta is regarded as a fixed class label. As a result, the problem of online disaster knowledge extraction is transformed into a problem of text multi-classification. The disaster social media text is divided into related event-meta classes by the multi-classification algorithm. However, the disaster social media data reveal that one flood disaster social media instance does not correspond to only one class (i.e., an event-meta). For example, a flood microblog instance may be related to medical and material needs and mentions the information about building damage. Therefore, the problem of online disaster knowledge extraction based on event-meta framework is a problem of multi-label learning. After this classification process, a flood social media text may have one or more event-meta tags.

The multi-label classification of flood microblogging is defined as follows:

Definition 1 The feature attribute set of each flood microblog is defined as $X \subset R^d$, which is a set of d-dimensional space on the real number field (R). Thus, all of the flood microblogs could be presented as the vector $X = [x_1, \cdots, x_d]$ in the d-dimensional space. The meta label set of each flood microblog is defined as $L \subset R^d$, which is a set of λ-dimensional space on the real number field (R). Thus, all of the flood microblogs' meta label could be presented as the vector $L = [l_1, \cdots, l_\lambda]$ in the λ dimensional space.

In the multi-label classification process, each microblogging instance X could be marked as a label set Y. This label set Y is a subset of L, that is, $Y \subseteq L$. If Y is represented by the dichotomy vector method, $y = [y_1, \cdots y_\lambda]$. Therefore, if $y_j = 1$, the j-th meta label is related to the microblogging instance; if $y_j = 0$, the j-th meta label is not related to the microblogging instance. Two multi-label classification ideas are widely adopted,

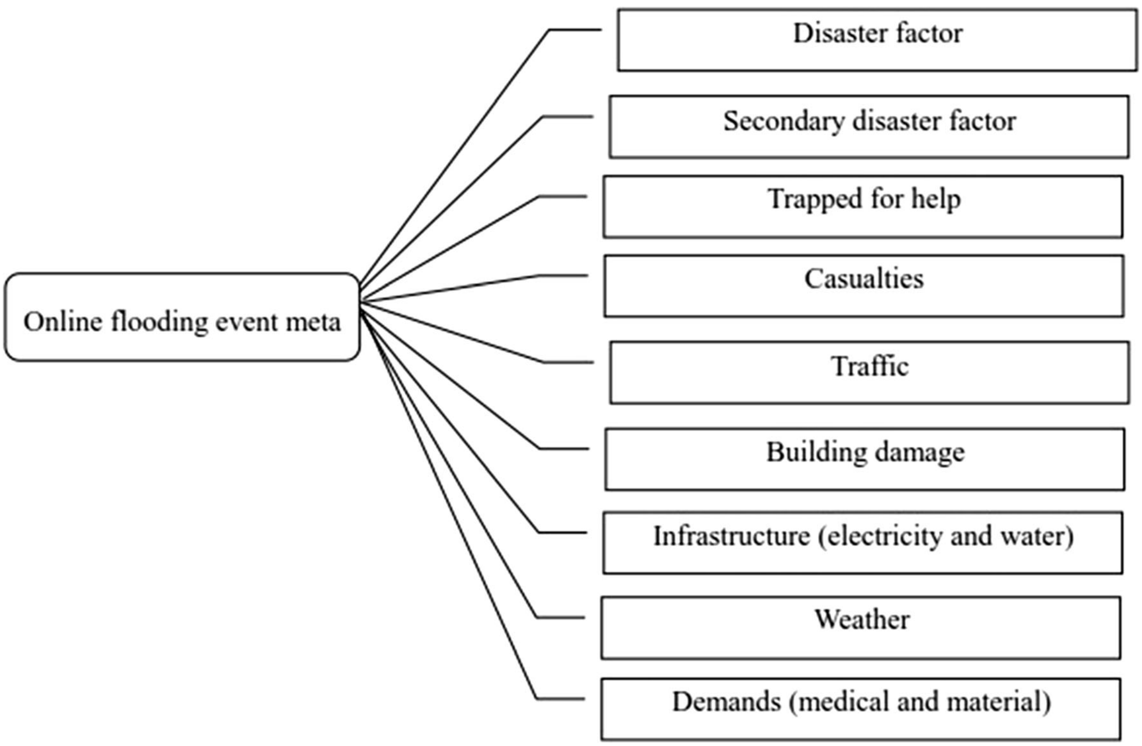


Fig. 3 Online flooding event-meta framework

namely, problem conversion and algorithm conversion. The problem conversion idea converts a multi-label problem into one or more single-label problems, classifies instances through traditional single-label classification methods, and converts the results into multi-labels, such as the random k-label set and the label power-set algorithm [56].

Algorithm conversion involves transforming the existing single-label learning algorithm to adapt to multi-label data. The basic idea of this method is to improve the traditional machine learning method such that it can solve the multi-label problem, such as Adaboost algorithm and multi-label KNN (MLKNN) algorithm [57]. The MLKNN algorithm has been widely used given its simple operation and good prediction effect. Therefore, this method is also used for multi-label classification in the present study.

3.3 Event-meta risk assessment of microblogging

In a flood disaster, the same user is in different disaster situations in different periods, and different users are in different disaster situations in the same period. Therefore, estimating the disaster situation that the user is in based on the flood-related microblogs posted by the user is necessary.

Based on the incremental multi-label classification model, flood-related microblogs could be converted into event-meta (class labels) for flood disaster situations. The risk of the event-meta to which flooding microbloggings belong can be estimated.

In the process of disaster and emergency disposal, according to the urgency and possible impact of an event, the risk of the event is generally divided into four levels, namely extraordinarily significant (Level I), significant (Level II), large (Level III), and general (Level IV). To serve emergency decision-making effectively, this study applies this grading standard to the risk assessment of disaster event elements. For flood disasters, nine event-metas are identified based on professional opinions and historical knowledge. Each event-meta is divided into four levels.

In order to find the appropriate event-metas, we first summarized the information needs of official flood disaster emergency response organizations (such as the emergency management department, the water resources department, the meteorological bureau, the natural resources department, etc.) and social organizations (such as the Red Cross department). Then, based on the researchers' outputs before [3, 5, 10, 17] and information requirements of and emergency response organizations, 20 event elements were drafted. Next, driven by the Delphi method, the opinions of authoritative 10 experts randomly selected were investigated and summarized. In the end, nine event-metas are identified finally.

For any online flood social media microblog x_i, its event-meta label vector $Y_i = \left(y_i^1, \cdots, y_i^j, \cdots, y_i^9\right)$ is obtained using the multi-label classification illustrated in Sect. 3.2. If $y_i^j = 1$, the risk assessment process of the event-meta will be implemented, and the result of meta risk assessment will be introduced to replace the value y_i^j. That is y_i^j no longer takes all values as 1 but can take them as 1, 2, 3, and 4. If $y_i^j = 0$, the original assignment result will remain unchanged and y_i^j continue to be 0. Therefore, the meta label vector $Y_i = \left(y_i^1, \cdots, y_i^j, \cdots, y_i^9\right)$ of microblog x_i is converted into the meta risk levels vector $P_i = \left(p_i^1, \cdots, p_i^j, \cdots, p_i^9\right)$ using the above four-classification process, where p_i^j represents the risk assessment levels of the j-th meta of microblog x_i. If $p_i^j = 0$, no relationship exists between microblog x_i and j-th event-meta.

The problem of meta risk assessment of users has been transformed into a machine learning problem of multi-classification. Given a strict increasing relationship among the four risk levels, this study uses the multi-class logical ordered regression (MLR) model to construct a logistic regression of the four-classification process.

3.4 Disaster situation risk assessment of users

Event-metas of flood social media microblogging represent the flood disaster situation from different aspects. The event-meta risk assessment results also describe the risk level of the disaster situation where the social media users are located.

Therefore, an event-meta is used as the decision feature vector for disaster situation assessment, and the event-meta level evaluation results of a user's microblogging are used as the input of the user event-meta vector to construct the vector space of the user flood risk assessment feature. Then, the risk level of the disaster situation in which the user is located is regarded as the predicted class label. The problem of user situation risk assessment has been transformed into a machine learning problem of multi-classification.

The entire conversion process is sorted as follows:

(1) The microblogs related to the flood disaster situation posted by users on a social media platform are labeled to identify the relevant event elements for users to post the information.
(2) The microblog texts about the flood are used as input data, the risk level of related event-metas are used as classification labels, and the related event element risk level of flood social media text is predicted using a four-classification algorithm.
(3) The event-metas and the risk levels of flood microblogging are used as classification labels, and the users' disaster situation risk levels are predicted using the four-classification algorithm.

The MLR algorithm is selected to assess users' disaster situation risk levels.

3.5 Regional flood disaster risk assessment

Weibo is a popular Chinese social media platform in China. After the flood disaster occurs, keyword queries can be carried out on the Weibo "Advanced Search" page to crawl microblogs from flood-affected areas. These crawled Weibo messages carry the users' "location city" data, which helps in clustering flood situation microblogs from the same city. Based on these datasets, the regional event-meta risk levels and flood situation risk levels could be evaluated at the urban-level scale. Moreover, users that post on social media with their GPS data using mobile phones are increasing. The precise GPS data facilitate the evaluation of the regional event-meta risk levels and flood situation risk levels at a smaller scale, such as township- and street-level scales.

Once a flood disaster occurs, massive situation microblogs immediately emerge on social media platforms. Therefore, the regional risk assessment needs to cluster microblogs into their regional group to address the challenges of large-scale text flow data. At a city-level scale, information from different users in the same city can be gathered based on the profile page information of Weibo users. At the township-level and street-level scales, microblogs from different users in different small-scale areas can be gathered based on the GPS data (i.e., latitude and longitude) and density clustering methods (e.g., DBSCAN).

The event-meta risk level of each flood disaster microblog has been evaluated. However, different users have different perceptions of disaster situations. The regional disaster risk level needs to consider the situations of different users comprehensively and remove noise data reasonably. To assess quickly the meta risk levels of a flood disaster event in a region, all of the microblogs from the region are gathered, and their event-meta risk levels are connected by column (meta name) to the regional event-meta vector space. The event-meta of the region is then obtained by identifying the mode of each column. The details of calculating are shown in Fig. 4. Here, x_i refers to the ith Weibo instance, p_i^j means the j-th event-meta risk levels of the ith Weibo instance, and $P_{w_t}^j$ is the j-th event-meta risk levels of region W_t.

Following the above research ideas, after clustering the flood disaster microblogs in each region, the event-meta

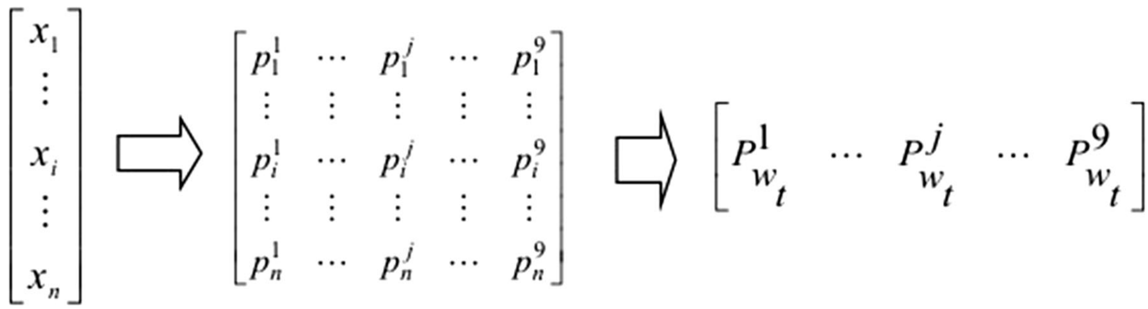


Fig. 4 Regional disaster event-meta calculated process

risk levels of each region can be calculated from the event-meta risk levels of the flood disaster microblogs in the region. Moreover, the event-meta risk levels of each region are used as an input vector, and the MLR flood risk evaluation model is used in the regional disaster situation risk evaluation. The evaluation process is shown in Fig. 5.

4 Case study

To verify the effectiveness of the above quantitative assessment model of flood disasters proposed in this research, the Yuyao Flood (2013, Zhejiang) is studied as a case.

4.1 Data set

Typhoon "Fitow" landed on the territory of Fujian Province, China at 9 am on October 7, 2013. Zhejiang Province experienced heavy rainfall, which has been rare for 100 years. Most of the main urban area of Yuyao City was submerged, and the transportation system was paralyzed (https://baike.so.com/doc/7188806-7412948.html).

Jisuke crawler is used to crawl data by keyword search on the "Advanced Search" page of Sina Weibo. Taking into account the complexity of Chinese expression, more than ten closely related words, including rain, big rain, heavy rain, typhoon, water, big water, flood, heavy floods, flooding, submerged, flooded, etc., are used as query keywords. In addition, according to the influenced scope of the research object (Yuyao Flood), the information geographic source of the advanced query port is set to Zhejiang Province. From 0:00 on October 6, 2013 to 0:00 on October 12, 2013 (120 h), 28,266 Weibo messages containing disaster keywords are obtained from users located in Zhejiang Province. After deduplication and denoising preprocessing, a binary classification support vector machine algorithm is used to identify the Weibo messages about the flood. The performance and stability of this algorithm have been confirmed by a series of previous studies [58]. Finally, 12,236 relevant Weibo messages about the Yuyao flood are screened and included in the data set of the case study. It

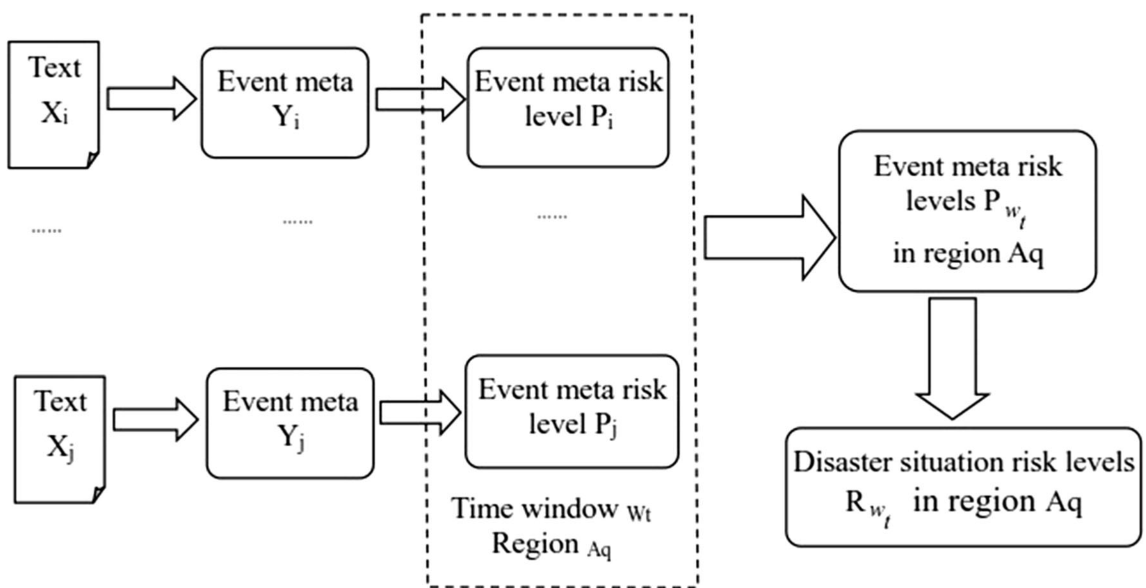


Fig. 5 Regional disaster event-meta and disaster situation risk assessment method in real-time

should be noted that because we focus on the first-hand information of witnesses, news data are treated as noise, but the witness responses under news are collected.

4.2 Experiment

First, based on the event-meta framework of flooding disasters, each Weibo message in the Yuyao flood dataset is classified by the multi-label classification algorithm. According to the labeling results, each flooding Weibo message from the affected area (i.e., Zhejiang Province) is represented by the vector with nine event-meta features, and each event-meta feature is represented by the Bourg vector (0 or 1). The random selected four Weibo messages instances, that have been translated into English, and their output results of after multi-label classification are shown in Tables 1 and 2.

Moreover, according to the multi-label classification results, each event-meta of the four microblogs corresponds to a one-dimensional Bourg feature value (i.e., 0 if related and 0 if unrelated). Similarly, each flood disaster microblog in the Yuyao flood data set is characterized as a nine-dimensional event-meta vector. The Bourg vector representation of the event-meta features of the above four flood disaster Weibo messages is shown in Table 3.

The event-meta risk levels of each flood disaster microblog in the Yuyao flood data set are classified into four classes. Thus, all of the flood situation microblogs are further characterized as a nine-dimensional vector that is composed of event-meta risk levels. The vector representation of the event-meta risk level of the above four flood disaster Weibo messages is shown in Table 4. Moreover, by considering the event-meta risk levels as a feature vector input, each microblog sample of the Yuyao flood dataset is classified into four flood situation risk levels. The situation risk levels of the above four flood disaster Weibo messages are shown in Table 4.

4.3 Results and discussion

To test the effectiveness of the proposed four-classification assessment model of the event-meta risk and the user's disaster situation risk, 10% of the samples for manual annotation from the Yuyao flood dataset are randomly selected. Then, the multi-class support vector machine (MSVM) and MLR algorithms are used to classify the labeled sample set. The macro-average F1 value and micro-average F1 value of the two four-classification algorithms can be obtained separately using the 10-dimensional cross test. In the evaluation process of the multi-classification models, in order to comprehensively consider the classification results of multiple categories, two evaluation indicators, micro-averaging and macro-averaging, are widely used. Micro-averaging and macro-averaging are obtained by different averaging methods. Micro-average index requires that the evaluation index value (F1) be calculated for each individual category first, and then take the arithmetic average of the evaluation index values of all categories as the average index (F1). Macro-average index requires the establishment of a confusion matrix for all data instances in the data set (statistics without classification), and then calculate the corresponding evaluation index Value (F1).

The comparison test results showed the following:

(1) During the four event-meta risk assessment process, the macro-average F1 value of the MSVM classification algorithm is 0.788, the micro-average F1 value of the MSVM classification algorithm is 0.754,

Table 1 Examples of four Weibo instances

Weibo instances	Text of each Weibo instance
[1]	Brothers, I am a refugee. Flooding, water and power outages, and I will soon lose contact with the outside world @clj777777 @Xiaolizi Eluting @Tingtingbao Diary @ Little Fish Who Loves Seafood @Xujinlong- I am:|Yangmingxi Road
[2]	Besieged by the flood for two days and two nights, the water has not diminished or subsided, and now it has reached the thighs, and the waist and neck are flooded with relatives. At present, the low-voltage water supply is cut off, and the mobile phone signal is sometimes not available. The mobile phone was out of power, and we lost contact with the outside world when we were trapped at home. No one came to inform the rescue. We did not tell us what the situation of the trapped residents is now and what will happen next. I am at:|Beibinjiang Road
[3]	Emergency call for help, our company was interrupted due to flooding, and there was no water, electricity or food. More than 30 employees requested food assistance. The address is No. 3 Jinxing Road, Yuyao, 15306662255(Mr. Chen). I'm at:|Yu-Ci connection line
[4]	Xicheng 12 courtyard has been connected to water and electricity, but the underground garage is closed with water. Pumping is underway. Lianfeng Middle Road to Market Port. The flood has receded. Many vehicles can pass slowly. Vehicles and pedestrians pay attention to safety@ Modern Gold News @ Ningbo Evening News @NBTV-2 to talk about what I am here:	Park Road

Table 2 Multi-label classification outputs of four Weibo instances

Weibo instances	Multi-label classification outputs of each Weibo instance
[1]	$\{y_1, y_3, y_7\}$
[2]	$\{y_1, y_3, y_7\}$
[3]	$\{y_1, y_3, y_5, y_7, y_9\}$
[4]	$\{y_1, y_5, y_7\}$

Table 3 Representation of Event-Meta Features for Disaster Weibo Text Examples

Weibo instances	Bourg vector representation of event-meta features
[1]	{1, 0, 1, 0, 0, 0, 1, 0, 0}
[2]	{1, 0, 1, 0, 0, 0, 1, 0, 0}
[3]	{1, 0, 1, 0, 1, 0, 1, 0, 1}
[4]	{1, 0, 0, 0, 1, 0, 1, 0, 0}

Table 4 Multi-evaluation input and output examples of flood situation risk levels

Weibo instances	Event-meta risk levels vector representation	Flood situation risk levels
[1]	{4, 0, 3, 0, 0, 0, 3, 0, 0}	4
[2]	{4, 0, 4, 0, 0, 0, 4, 0, 0}	4
[3]	{4, 0, 4, 0, 4, 0, 4, 0, 3}	4
[4]	{2, 0, 0, 0, 2, 0, 1, 0, 0}	2

and the macro-average F1 value of the MLR classification algorithm is 0.814, and the micro-average F1 value of the MLR classification algorithm is 0.797.

(2) During the four users' disaster situation risk assessment process, the macro-average F1 value of the MSVM classification algorithm is 0.896, the micro-average F1 value of the MSVM classification algorithm is 0.872, the macro-average F1 value of the MLR classification algorithm is 0.927, and the micro-average F1 value of the MLR classification algorithm is 0.920.

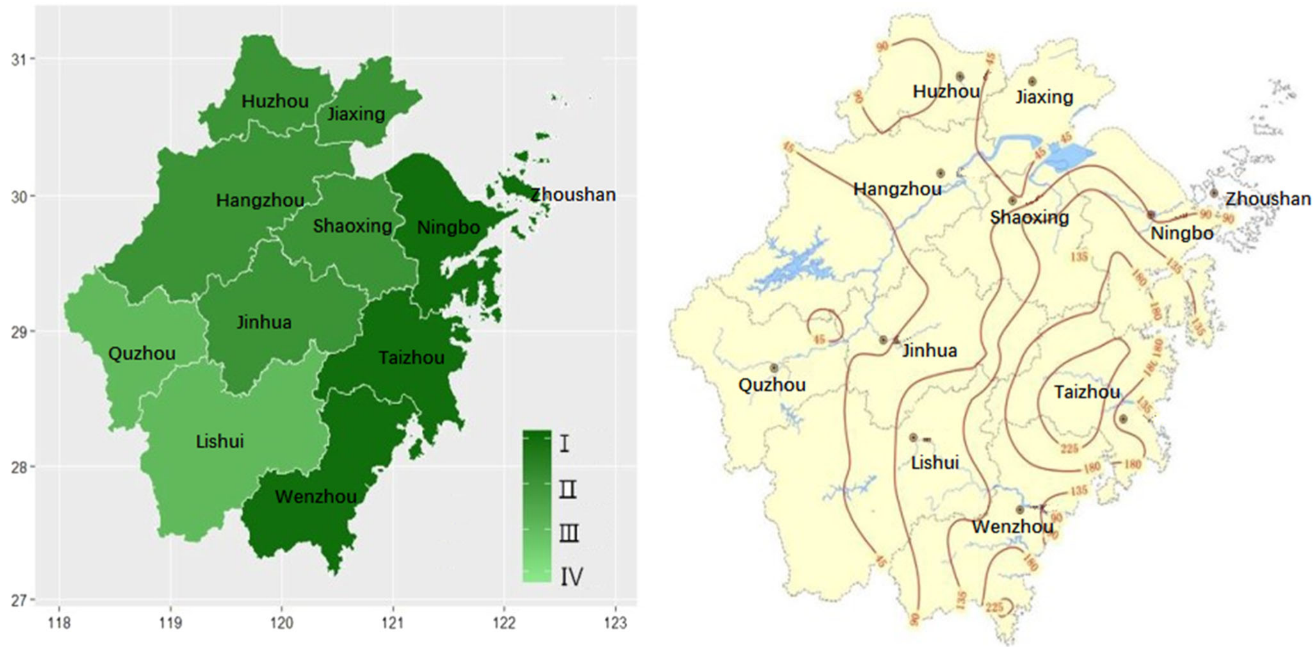
The comparison test results show that the MLR algorithm has better performance than the MSVM classification algorithm in the event-meta risk assessment and the user disaster situation risk assessment. Therefore, choosing the MLR algorithm for rapid assessment of the flood microblogs is reasonable.

Next, to provide an intuitive decision-making reference for the disaster emergency management process, all the Weibo samples in the Yuyao flood data set are gathered in different groups according to the time window w_t (e.g., 24 h) and the source region (city) A_q. Then, in different time windows, the event-meta risk in different regions is calculated by the mode method. Thus, the regional risk of the flood disaster situation is classified into four levels.

At present, authoritative statistical data on flood risks are lacking. The flood risk is closely related to many influencing factors that are difficult to be quantified, such as rainfall, tides, poor drainage in local areas, and reservoir flood discharge. The influencing factors of the Yuyao flood are relatively complex. However, the large-scale precipitation brought by Typhoon "Fitow" is the direct cause of this disaster and the most important factor that influence the flood risk in Yuyao. Thus, the available and quantifiable regional precipitation data are used to analyze the effectiveness of this study's novel assessment model.

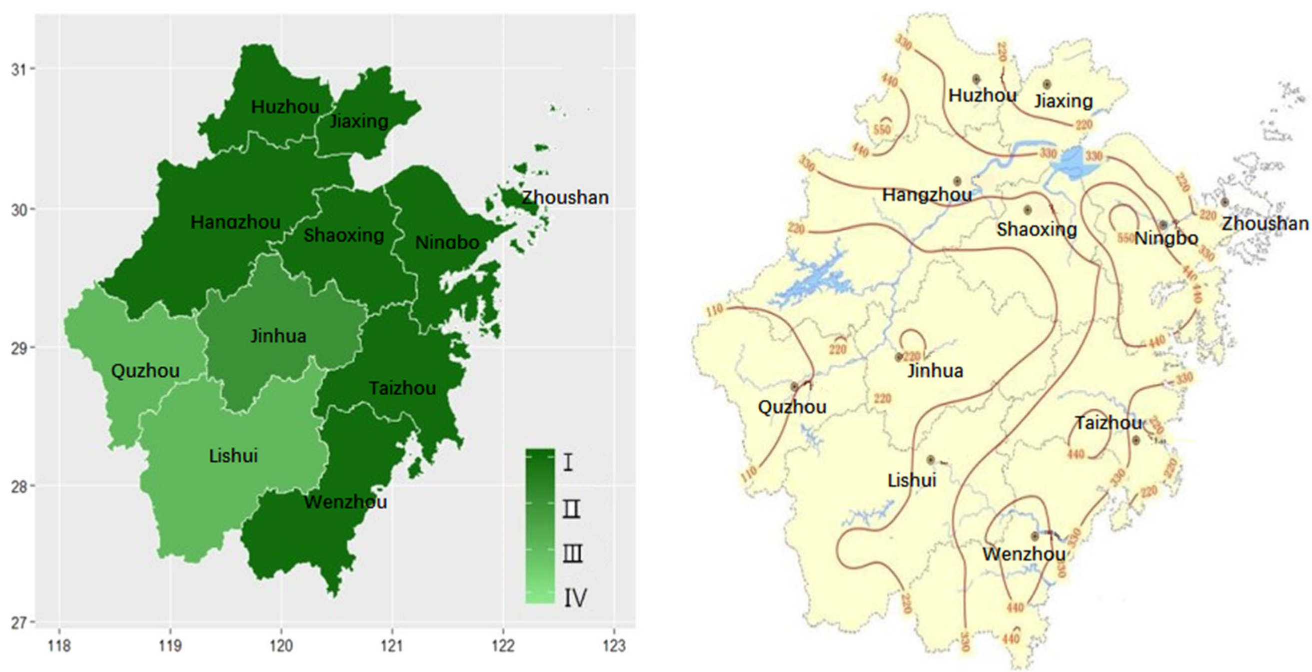
After analyzing the precipitation data of the four observatories in Zhejiang Province, the duration of heavy rainfall brought by "Fette" is found to be from October 6–8, 2013.

Figures 6 and 7 show the maps of the regional flood risk assessment results of cities obtained by the proposed rapid assessment based on the flooding situation Weibo messages from Zhejiang Province in the first time window w_1 and the second time window w_2. To confirm the validity of the evaluation results, the maps of precipitation in the same time window are listed in Figs. 6 and 7, and the precipitation data come from the China Meteorological



(a) Flood risk assessment levels in w_1 **(b)** Cumulative precipitation in w_1

Fig. 6 Disaster situation risk levels and accumulative precipitation in w_1



(a) Flood risk assessment levels in w_2 **(b)** Cumulative precipitation in w_2

Fig. 7 Disaster situation risk levels and accumulative precipitation in w_2

Administration Meteorological Data Center. Figures 6 and 7 show that the regional disaster risk level released by the victims is generally consistent with the distribution of precipitation brought by "Fette" in w_1 and w_2. This result shows that, within these two time windows, the flood disasters occurred, and precipitation is an important factor for flood disaster risk in the affected region.

Figure 8 shows the disaster situation risk levels in the w_3 and w_4 of different cities in Zhejiang Province. In w_3, as "Fette" gradually passed through and left Zhejiang, precipitation in various regions (cities) in Zhejiang Province decreased significantly, and rain stopped in many cities. However, as shown in Fig. 8, affected by drainage and other factors, the flood disaster situation risk levels reported by victims in Hangzhou, Ningbo, and Huzhou in the time window w_3 were serious and could not be effectively alleviated. In w_4, the flood disaster situation in Hangzhou and Huzhou were controlled, but the flood disaster situation in Ningbo continued.

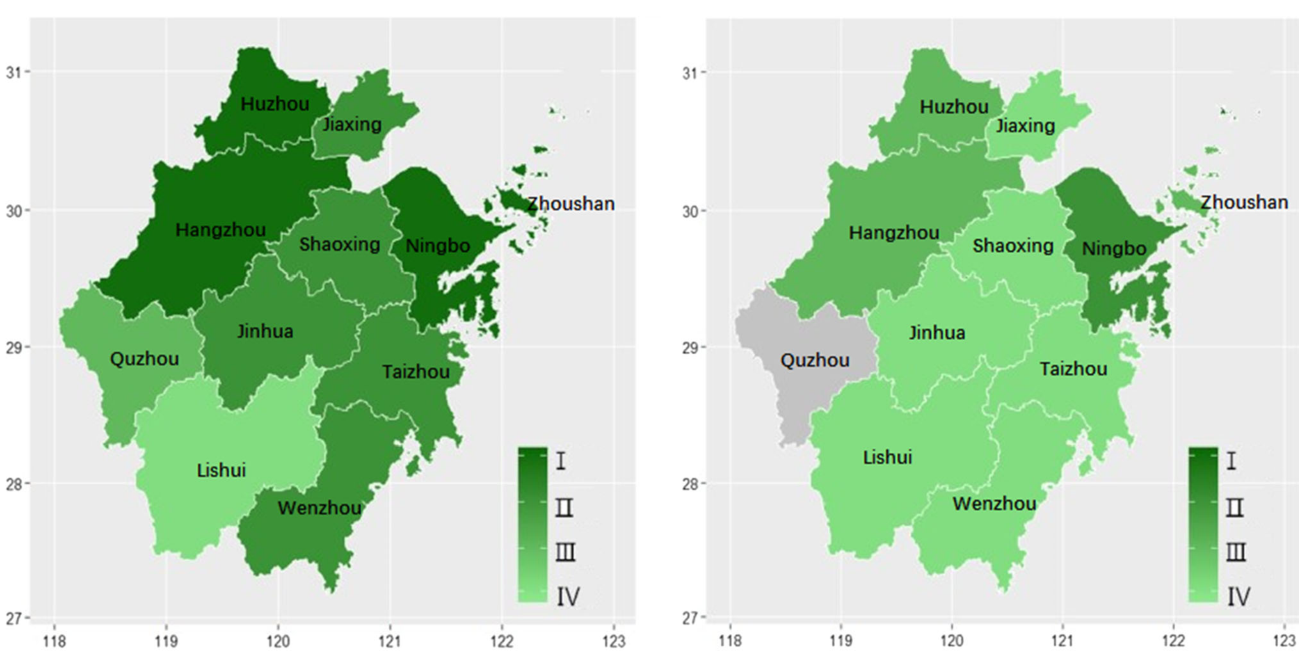
Figure 9 shows the flood risk level of the eleven cities in Zhejiang Province and the flood risk level of each county in Ningbo City in w_5 based on the situation microblog evaluation of the victims' feedback. Although in Ningbo City, the flooding situation in most counties has been greatly alleviated, the risk level in Yuyao remains relatively high in w_5.

Yuyao is a county-level city managed by Ningbo City. In a short period of time, Typhoon Fitow brought 750 million cubic meters of precipitation to Yuyao, which is equivalent to directly spilling water from 68 West Lakes onto the land of Yuyao. In addition, Yuyao's geographic location is similar to the "bottom of the pot." A series of factors, including the astronomical tide, the continuous flooding in the upstream, and the high tide level in the downstream, make it difficult for Yuyao's floods to be discharged in time, and the water level remains high for a long time. This is the main reason the flood in Zhejiang Province caused by Typhoon Fitow was finally named to be Yuyao Flood. The data analysis results of the case study could effectively support the reality of Yuyao Flood.

5 Conclusion

With the continuous acceleration of global warming and urbanization, casualties and property losses caused by frequent flood disasters have gradually increased. Determining how to understand and assess the danger of flooding quickly has attracted the attention of many researchers. The following are expected in the process of evaluating the risk of flood disasters. Meteorological data (e.g., precipitation) can be obtained in time. Geographic elevation data (RS) are available, but a lag period is anticipated. Many special disaster-pregnancy environment data, such as drainage pipes, are blocked or backfilled. The location of the victims is difficult to obtain. The lag period is long. The widespread use of mobile phones and social media tools has provided the possibility and convenience for flood disaster risk assessment. Disaster streaming on social media is helpful to supplement the deficiencies of real-time data acquisition.

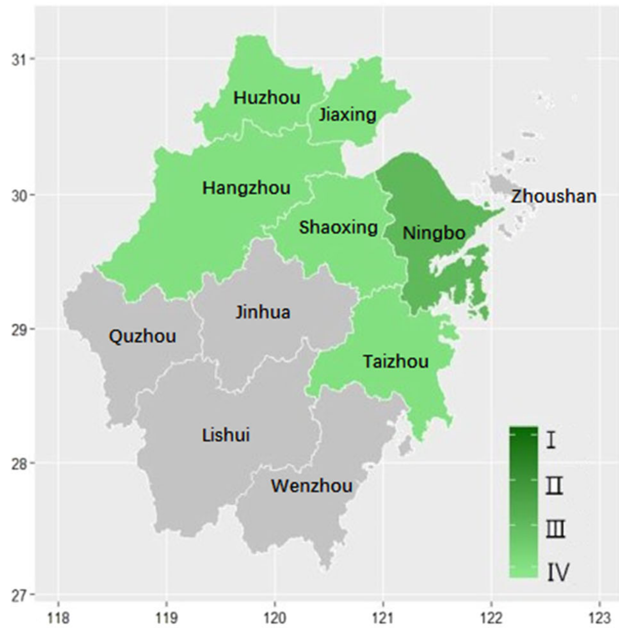
To achieve the objectives of effective flood disaster information extraction and the rapid assessment of flood



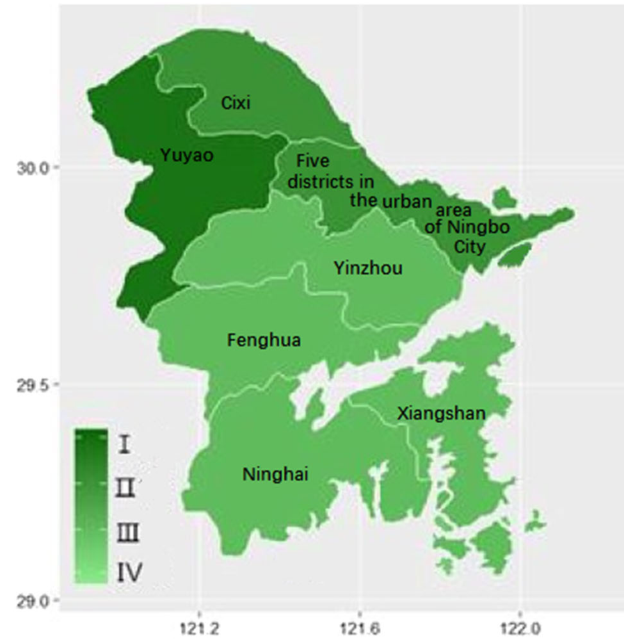
(a) Flood risk assessment levels in w_3

(b) Flood risk assessment levels in w_4

Fig. 8 Disaster situation risk levels in w_3 and w_4



(a) Zhejiang flood risk assessment levels in w_5



(b) Ningbo flood risk assessment levels in w_5

Fig. 9 Disaster situation risk levels in w_5 by different scales

risk levels, this study establishes a flood disaster event-meta framework based on the needs of disaster emergency response decision-making. According to the text data mining algorithm, event-metas are used as class labels, and a social media post on the flood is classified into multiple labels to identify the event-meta to which the online flood microblog belongs. Then, the flood microblog texts that belong to a certain event-meta are used as input vector, and the logistic regression ordered multi-classification algorithm is applied to carry out the four-classification process (four levels) of event-meta risk levels to the flood microblog. As a result, the event-meta class labels of the flood microblogs are converted into event-meta risk level labels. Next, the event element of the flood microblogs is used as a decision feature to construct a feature space. The evaluation result of the event element rank is used as the input vector of the feature space, which is used to establish a rapid risk assessment model of the flood microblog's disaster situation. Finally, the mode of the risk level of event elements in the region is used to establish the feature space of the event risk level of the region and identify the flood disaster event elements and their risk levels in the region. The regional flood disaster risk level has been quickly evaluated.

The results of the Yuyao flood case study show that the MLR classification algorithm used in the rapid quantitative evaluation model established in this study is efficient. The macro-average F1 and micro-average F1 of MLR are higher than those of MSVM. Moreover, the interpretability and validity of the model evaluation results are confirmed by the precipitation data in Zhejiang Province. Furthermore, this assessment method can be applied to small-scale areas, such as a county (district), township (town), and street (village) flood hazard assessment, which would effectively improve the traditional disaster information collection approach.

The risk of flood disasters is related to precipitation and other meteorological conditions and subject to complex comprehensive influencing factors, such as regional terrain and drainage system. Therefore, in the process of flood disasters management, evaluating the disaster risk in different regions in real-time is a complex and difficult task. The proposed social media platform-based disaster risk assessment method for flood disaster areas fully uses the information posted by the victims and integrates objective impact factors by using the intuitive description of the disaster experience for assessment. Moreover, the proposed regional disaster risk calculation method further optimizes the impact of fuzzy texts and extreme descriptions of disaster victims on the final evaluation results and can provide effective decision-making references for the flood disaster emergency management process. In addition, the method proposed in this paper can also be used for social media data in English and other languages in disaster situation. For the future research in other languages and other types of disasters, the most important work is the identification of event-metas and the selection of text classification algorithms.

This is necessary and important, that is, a new framework proposed in this paper served the quantitative analysis of self-generated text descriptions for flood disaster witnesses. However, the witnesses post pictures and short videos with their texts together usually. Unfortunately, the pictures and short videos carried by Weibo data did not be considered and processed currently. In the future, we will work on the fusion of video, pictures and text, and try to combine social media pictures with 3S data for flood risk and loss assessment.

Funding This work was supported by the National Natural Science Foundation of China (Grant Nos. 71774041 and 71531013), Natural Science Foundation of Fujian Province (Grant No. 31185015), and Social Science Planning Project of Fujian Province (Grant No. FJ2018C008).

Compliance with ethical standards

Conflicts of interest The authors declare that they have no competing interests.

References

1. Manoj BS, Baker AH (2007) Communication challenges in emergency response. Commun ACM 50:51–53
2. Restrepo-Estrada C, de Andrade SC, Abe N et al (2018) Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. Comput Geosci 111:148–158
3. Scott KK, Errett NA (2018) Content, accessibility, and dissemination of disaster information via social media during the 2016 Louisiana floods. J Public Health Manage Practice 24:370–379
4. Cheng J, Sun AR, Hu D et al (2011) An information diffusion based recommendation framework for micro-blogging. J Assoc Inf Syst 12:45–73
5. Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int J Inf Manage 31:6–13
6. Latapi Agudelo MA, Johannsdottir L, Davidsdottir BA (2019) literature review of the history and evolution of corporate social responsibility. Int J Corporate Social Respons 4(1):1–23
7. Standing S, Standing C (2018) The ethical use of crowdsourcing. Business Ethics A Europ Rev 21(1):72–80
8. Kraljevic D, Lackovic K, Sojo R (2020) The information—communication process in a business with outsourcing for the maintenance of a complex technical system. Tehnicki glasnik 14(2):194–200
9. Mccallum I, Liu W, See L et al (2016) Technologies to support community flood disaster risk reduction. Int J Disaster Risk Sci 7(2):198–204
10. Bunce S, Partridge H, Davis K (2012) Exploring information experience using social media during the 2011 queensland floods: a pilot study. Aust Library J 61(1):34–45
11. Yin J, Karimi S, Robinson B, et al (2012) ESA: emergency situation awareness via Microbloggers. In: Proceedings of the 21st ACM international conference on Information and knowledge management, pp 2701–2703
12. Chowdhury SR, Imran M, Asghar M R et al (2013) Tweet4act: Using incident-specific profiles for classifying crisis-related messages. In: Proceedings of 10th international ISCRAM conference, pp 834–839
13. Rogstadius J, Vukovic M, Teixeira CA et al (2013) CrisisTracker: crowdsourced social media curation for disaster awareness. IBM J Res Dev 57:1–13
14. Robinson B, Bai H, Power R, Lin X G(2014) Developing a Sina Weibo incident monitor for disasters. In: Proceedings of the Australasian language technology association workshop, pp 59–68
15. Acar A, Muraki Y (2011) Twitter for crisis communication: lessons learned from Japan's tsunami disaster. Int J Web Based Commun 7(3):392–402
16. Gao H, Barbier G, Goolsby R (2011) Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell Syst 26(3):10–14
17. Li J, Rao HR (2010) Twitter as a rapid response news service: an exploration in the context of the 2008 China earthquake. Electron J Inf Syst Develop Countries 42(1):1–22
18. Keim ME, Noji E (2011) Emergent use of social media: a new age of opportunity for disaster resilience. American Journal of Disaster Medicine 6(1):47–54
19. Kaufhold M, Reuter C (2016) The self-organization of digital volunteers across social media: the case of the 2013 European floods in Germany. J Homel Secur Emerg Manage 13(1):1–11
20. Ling CLM et al (2015) Ict-enabled community empowerment in crisis response: social media in Thailand flooding 2011. J Assoc Inf Syst 16(3):174–212
21. Alexander DE (2014) Social media in disaster risk reduction and crisis management. Sci Eng Ethics 20(3):717–733
22. Palen L (2008) Online social media in crisis events. Educause Quarterly 31(3):76–78
23. Thelwall M, Stuart D (2007) RUOK? Blogging communication technologies during crises. J Comput Mediat Commun 12(2):523–548
24. Crooks A, Croitoru A, Stefanidis A et al (2013) #Earthquake: twitter as a distributed sensor system. Trans GIS 17(1):124–147
25. Earle P, Guy M, Buckmaster R et al (2010) OMG earthquake! Can Twitter improve earthquake response? Seismol Res Lett 81(2):246–251
26. Allen RM (2012) Transforming earthquake detection? Science 335(6066):297–298
27. Feldman D, Contreras S, Karlin B et al (2016) Communicating flood risk: looking back and forward at traditional and social media outlets. Int J Disaster Risk Reduction 15:43–51
28. Velev D, Zlateva P (2012) Use of social media in natural disaster management. Int Proceed Econ Develop Res 39:41–45
29. Han XH, Wang JL, Bu K et al (2018) Progress in information acquisition of disaster events from web texts. J Geo-Inf Sci 20(8):1037–1046
30. Boulos MNK, Resch B, Crowley DN et al (2011) Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. Int J Health Geogr 10(1):67–71
31. Murakami A, Nasukawa T (2012) Tweeting about the tsunami?: mining twitter for information on the Tohoku earthquake and tsunami. In: Proceedings of the 21st international conference companion on world wide web, pp 709–710
32. Warner JE (2012) Social media to the rescue: the American red cross story. http://thesocialmediamonthly.com/social-media-to-the-rescue-the-american-red-cross-story/
33. Kim J, Hastak M (2018) Social network analysis: characteristics of online social networks after a disaster. Int J Inf Manage 38(1):86–96

34. Cervone G, Sava E, Huang Q et al (2016) Using twitter for tasking remote-sensing data collection and damage assessment: 2013 boulder flood case study. Int J Remote Sens 37(1–2):100–124
35. Singh YP, Yogesh K, Dwivedi NP et al (2019) Event classification and location prediction from tweets during disasters. Annals Oper Res 283(1):737–757
36. Ryan B (2013) Information seeking in a flood. Disaster Prevent Manage 22(3):229–242
37. Murthy D, Longwell SA (2013) Twitter and disasters: the uses of twitter during the 2010 Pakistan floods. Inf Commun Soc 16(6):837–855
38. De Albuquerque JP, Herfort B, Brenning A, Zipf A (2015) A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. Int J Geogr Inf Sci 29(4):667–689
39. Al-Saggaf Y, Simmons P (2015) Social media in Saudi Arabia: exploring its use during two natural disasters. Technol Forecast Soc Chang 95:3–15
40. Coz JL et al (2016) Crowdsourced data for flood hydrology: feedback from recent citizen science projects in Argentina, France and New Zealand. J Hydrol 541(10):766–777
41. Arthur R, Boulton CA, Shotton H et al (2018) Social sensing of floods in the UK. PLoS ONE 13(1):1–24
42. Venganzones-Bodon M (2019) Machine learning challenges in big data era. DYNA 94(5):478–479
43. Chaw HT, Kamolphiwong S, Wongsritrang K (2019) Sleep apnea detection using deep learning. Tehnicki glasnik 13(4):261–266
44. Vieweg S, Hughes A L, Starbird K et al (2010) Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1079–1088
45. Shklovski I, Burke M, Kiesler S (2010) Technology adoption and use in the aftermath of Hurricane Katrina in New Orleans. Am Behav Sci 53(8):1228–1246
46. Neubaum G, Rösner L, Rosenthal-von der Pütten AM et al (2014) Psychosocial functions of social media usage in a disaster situation: a multi-methodological approach. Comput Hum Behav 34:28–38
47. Fohringer J, Dransch D, Kreibich H, Schroter K (2015) Social media as an information source for rapid flood inundation mapping. Nat Hazards Earth Syst Sci 3(7):4231–4264
48. Smith L, Liang Q, James P et al (2017) Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. J Flood Risk Manage 10(3):370–380
49. Li Z, Wang C, Emrich CT et al (2018) A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. Cartography Geogr Inf Sci 45(2):97–110
50. Brenden J, Jurjen W, Beatriz R et al (2015) Early flood detection for rapid humanitarian response: harnessing near real-time satellite and twitter signals. ISPRS Int J Geo-Inf 4(4):2246–2266
51. Feng Y, Sester M (2018) Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos. ISPRS Int J Geo-inf 7(2):39
52. Artzai P, Aitor AG, Unai I et al (2020) Why deep learning performs better than classical machine learning? DYNA 95(2):119–122
53. Liang CY, Zhang GF, Wang MF et al (2018) Assessing the effectiveness of social media data in mapping the distribution of typhoon disasters. J Geo-Inf Sci 20(6):807–816
54. Chen Z, Gao T, Luo NX et al (2017) Empirical discussion on relation between realistic disasters and social media data. Sci Survey Map 42(8):44–48
55. Bai H, Yu H L, Yu G et al (2020) Automatic information extraction and quantitative risk assessment of flood disaster microblog based on Word2vec and Multi-class Logistic Regression. In: Proceeding of WorldCist20, pp 1–5
56. Tsoumakas G, Katakis I (2007) Multi-label classification: an overview. Int J Data Warehouse Min 3:1–13
57. Zhang M, Zhou Z (2007) ML-kNN: a lazy approach to multi-label learning. Pattern Recognit 40(7):2038–2048
58. Bai H, Yu G (2016) A Weibo-based approach to disaster informatics: incidents monitor in post-disaster situation via Weibo text negative sentiment analysis. Nat Hazards 83(2):1177–1196